

Conceptos básicos de estadística

Sesión 02

Ing. Gómez Marín, Jaime¹

Módulo 2 : Estadística y Visualización de Datos con Python
Departamento de TdG

September 2019



- Introducción
- Estadísticas
- Conclusiones
- Bibliografía



En esta sesión se aprenderá los conceptos básicos de probabilidades y estadística , se usa Python para el desarrollo de los ejercicios



Son Estadísticos o valores numéricos que muestra el centro de los datos, entre los que tenemos:

- media aritmética
- mediana
- moda



Sea X_1, X_2, \dots, X_n elementos de una muestra, se define como media aritmética o promedio a:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

```
1 # Media
2 def cal_media():
3     print("-----")
4     mean = df['Oro'].mean()
5     sumatoria = np.sum(df['Oro'])
6     nro_item = np.size(df['Oro'])
7     print("sumatoria ", sumatoria)
8     print("elementos ", nro_item)
9     print("media manual = ", sumatoria/nro_item)
10    print("media = ", mean)
```

Es un conjunto de n observaciones ordenadas de forma creciente, donde la mediana me se define como el valor central:

- Si n es impar, entonces, $me = X_{(n+1)/2}$
- Si n es par, entonces, $me = X_{n/2} + 1/2$



```
1 # Mediana
2 def cal_mediana():
3     print("-----")
4     mediana = df['Oro'].median()
5     nro_item = np.size(df['Oro'])
6     pos_mediana = round(nro_item/2)
7     print("pos_mediana = ", pos_mediana)
8     print("mediana manual = ",
9           df['Oro'][pos_mediana-1])
10    print("mediana = ", mediana)
```


Es el valor que tiene la mayor frecuencia

```
1 # Mode
2 def cal_mode():
3     print("-----")
4     print(df['Oro'].value_counts())
5     mode = df['Oro'].mode()
6     print("moda = " , mode)
```

Es un conjunto de n observaciones ordenadas de forma creciente, donde la mediana m_e se define como el valor central:

- Rango: es la diferencia entre el valor máximo y mínimo de los datos
- Cuártiles : La mediana de una distribución divide los datos en dos partes iguales, Cuando se dividen un conjunto ordenado de datos en cuatro partes iguales, los puntos de división se conocen como cuartiles y los representamos por Q_1 , Q_2 y Q_3 .
- Rango Intercuartílico (RI) : El rango entre cuartiles mide la variabilidad de la mitad central de los datos $Q_3 - Q_1$.



- Percentil : el k -ésimo percentil es un valor tal que el $k\%$ de los datos son menores o iguales que él, y el $(100 - k)\%$ restante son mayores o iguales que él
- Desviación estándar: es un promedio de las desviaciones individuales de cada observación con respecto a la media de una distribución.
- Varianza : Es el cuadrado de la desviación estándar

```
1  
2 # Percentiles  
3 def cal_percentiles():  
4     print("-----")  
5     tramos = [25,50,75]  
6     percentiles = np.percentile(df['Oro'],tramos)  
7     print("percentiles = ", percentiles )
```

```
1 # Varianza
2 def cal_varianza():
3     diff = df[['Oro']] - df['Oro'].mean()
4     print(type(diff))
5     diff2 = np.power(diff,2)
6     print(type(diff2))
7     nro_item = np.size(df['Oro'])
8     var_manual = (diff2.sum()/nro_item)['Oro']
9     #print(type(var_manual))
10    print("varianza manual = ", var_manual)
11
12    var = df['Oro'].var(ddof=0) # corrección de Bessel
13    print("varianza = ", var)
14    print("varianza = ", np.var(df['Oro']))
```

Desviación Estandar

```
1 # Desviacion Estandar
2 def cal_desviacion_estandar():
3     std_manual = np.sqrt(df['Oro'].var(ddof=0)) # co
4     print("desviacion estandar manual = ", std_manua
5     std = df['Oro'].std(ddof=0) # corrección de Bess
6     print("desviacion estandar = ", std)
```



En esta sesión se han tocado temas acerca de la estadística usando Pandas.



Fabio Nelli. Python Data Analytics with Pandas, NumPy, and Matplotlib, 2018.